VOCAL-INSTRUMENT SEPARATION
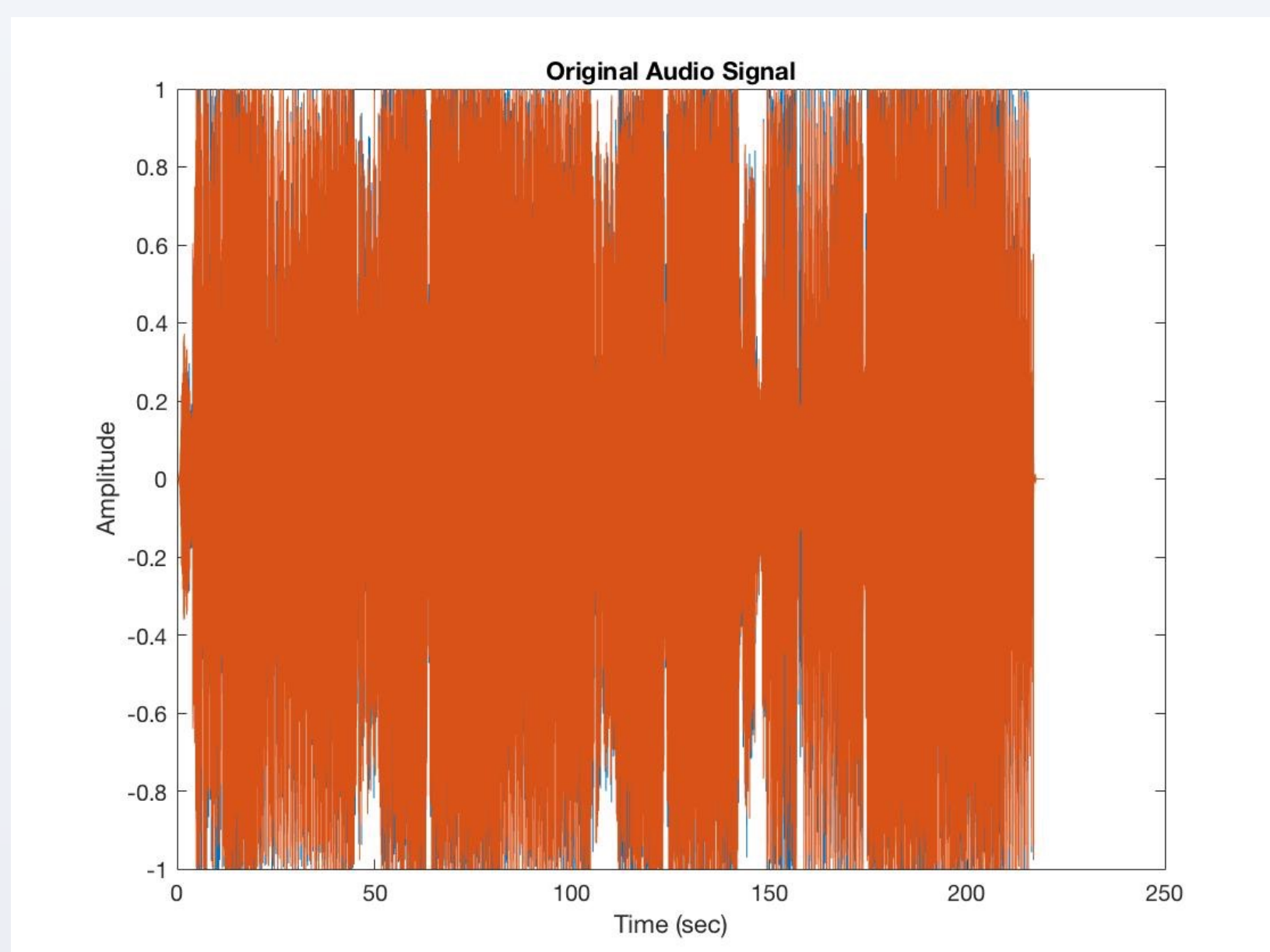
Dahyun Chung and Thomas Downey
University of Rochester Department of Audio and Music Engineering
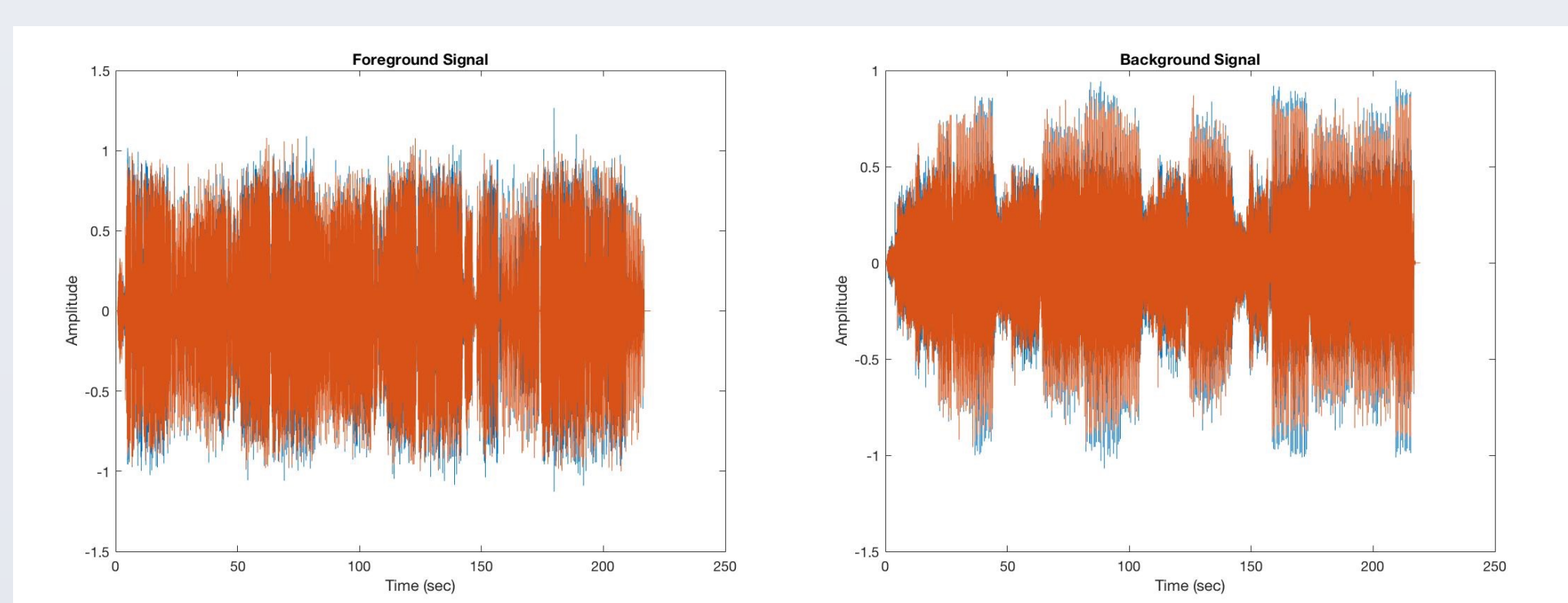
SUCCESSFUL IMPLEMENTATION OF REPET

When running REPET, the best separation of vocal and instrumental tracks in popular music came from music with a high male voice. Below are some visualizations of the separation of Aziatix's "Baby Let's Go":
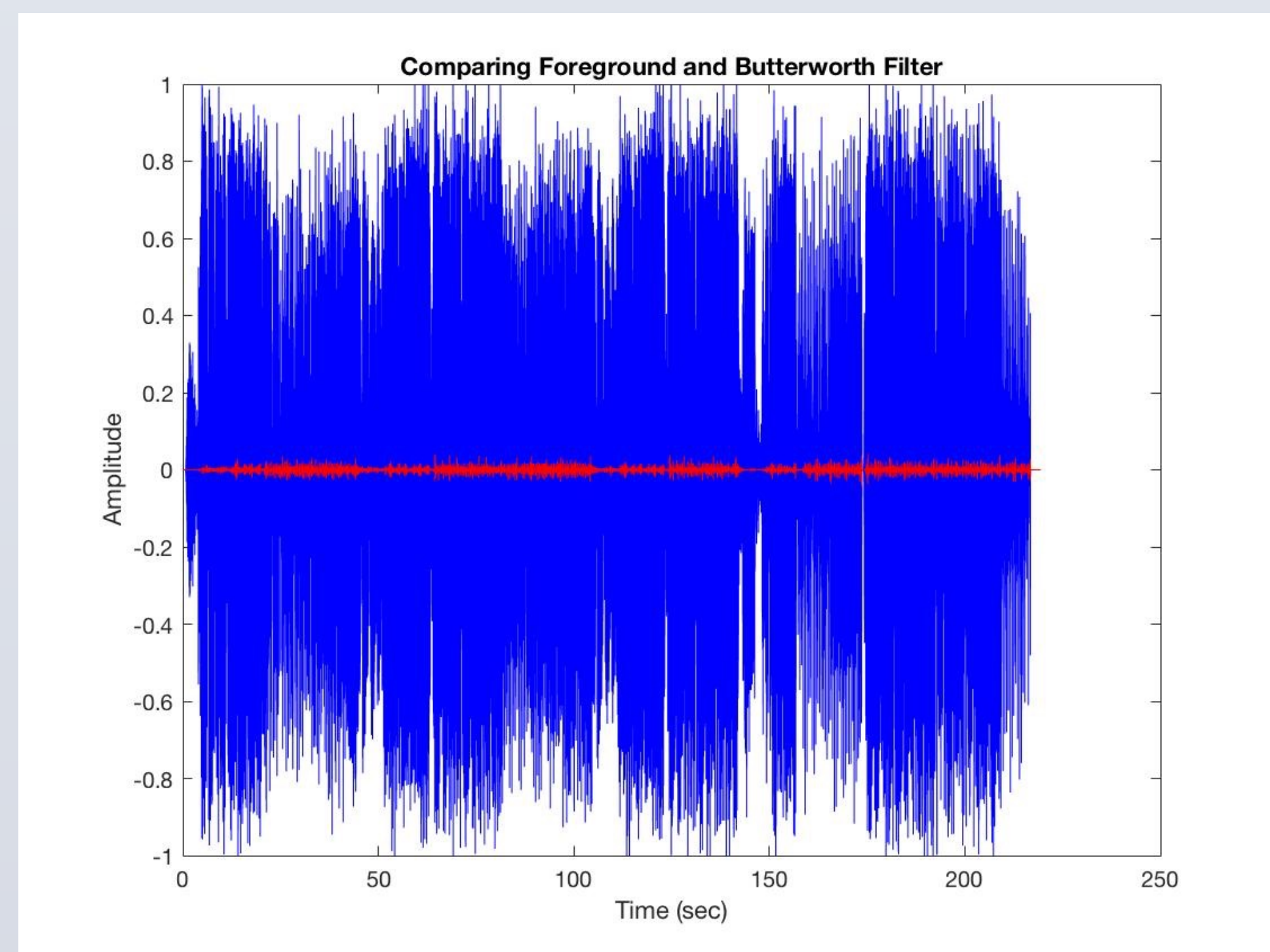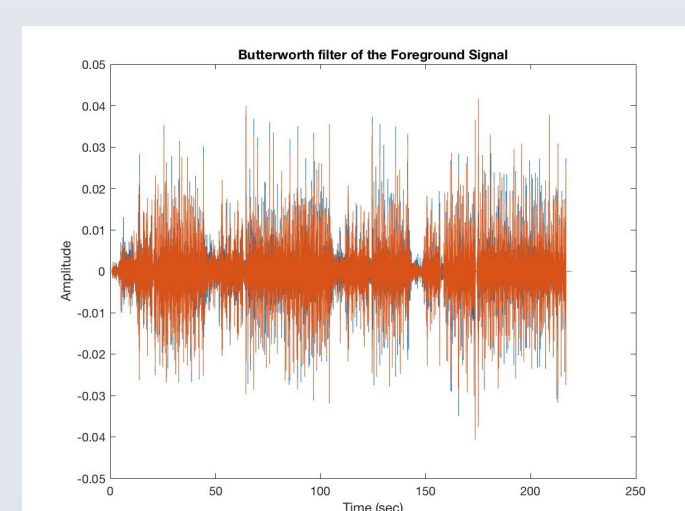

Original REPET

Vocal Melody   *Instrumental Background*

Butterworth Filter

PROJECT INTRODUCTION AND ABSTRACT

Current forms of popular music consist of two main layers: one or more lead vocalists singing a melody, and a collection of background instruments. There is considerable interest in giving consumers the ability to separate these two parts of the song, creating one vocal melodic track and one instrumental background track. There exists a free MATLAB program that is somewhat successful at filling this concept. However, distorted background tracks and the addition of high frequencies in separated tracks have make the program less than perfect. This project observes the successes and failures of the code and attempts to offer improvements to REPET. Our focus in on the accuracy of vocal melody extraction.
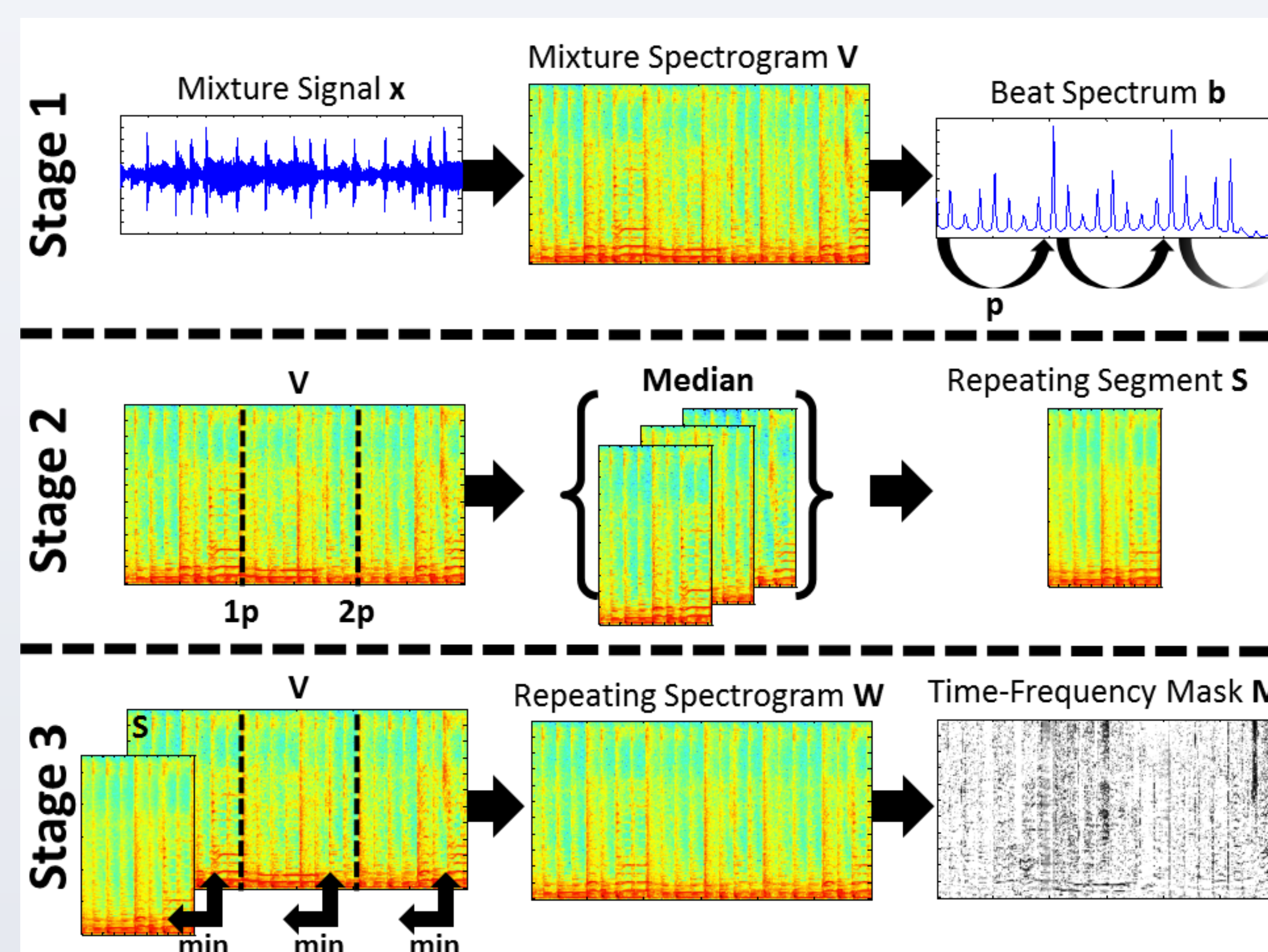
Figure: Overview of REPET, Zafar Rafii, zafarrafii.com

REPET: REpeating Pattern Extraction Technique (REPET) is a MATLAB code that is used to separate vocal melodies from instrumental backgrounds in popular music. Since background parts repeat more often than vocal melodies, the code finds repeating patterns in the music, attributes this to the background, and removes it from the melody. This method was designed by Zafar Rafii while pursuing his PhD. at Northwestern University. The code produces separated tracks with a quality that is comparable to expensive software and other high quality separation methods.

There are three stages in the REPET code:
1. A spectrogram of the input audio is created, and a repeating period is calculated from the beat spectrum.
2. Repeating patterns are identified from the spectrogram. Then the median of those repeating segments is calculated.
3. Repeating segments are subtracted from the original spectrogram.

CODE IMPROVEMENTS

The program REPET was executed multiple times for each song observed.
The separation of vocal tracks was improved most when we...

Changed the Cut-off Frequency:
In the original code, the high-pass cut-off frequency was set to 100 Hz. However, in most popular songs, vocalists sing consistently higher than 100 Hz. We increased the cutoff frequency to different values, depending on the nature of the vocals in the song. This improved the removal of drums and bass sounds from the vocal track.

Varied the Range of the Repeating Time Period:
The repeating time period sets the minimum and maximum length of the removed repeating section. In our trials, it was found that a small repeating period improved drum removal accuracy, and a large repeating period improved the removal of longer repeating harmonic sections.

Added a High-Pass Butterworth Filter:
In addition to the change in cutoff frequency in the original code, an additional high-pass Butterworth filter was applied to the vocal melody track. This filter was very helpful in completely reducing overall bass sound in vocal tracks. An additional low-pass filter was added as well to reduce the unnecessary high frequencies produced above 16 KHz.
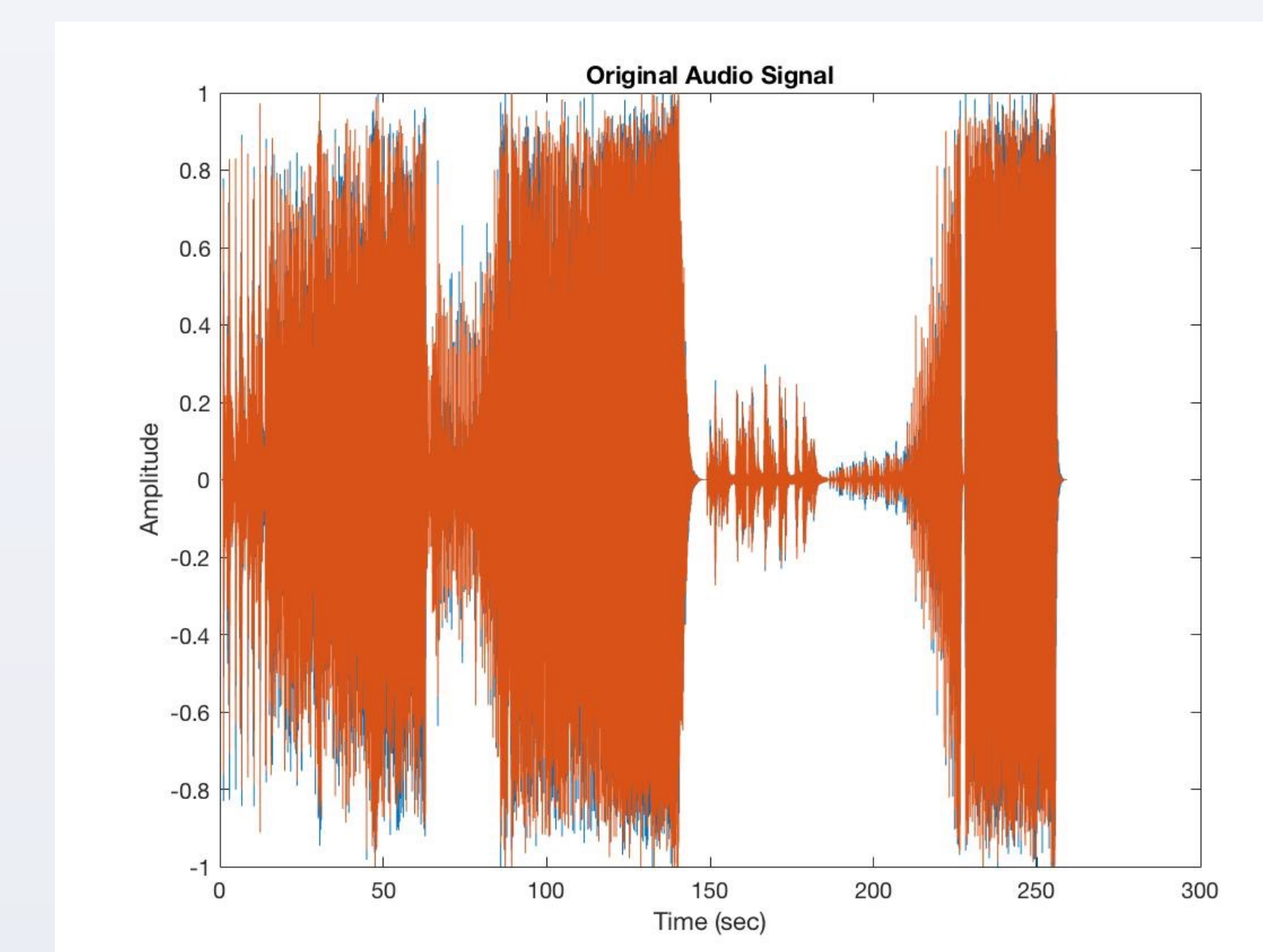
CONCLUSIONS AND FUTURE WORK

One conclusion to be made from this project is that, despite the multiple filters and the change in repeating time period, there is no perfect code that removes all vocal melodies from background instruments. While modifications for each specific song improved the clarity of the vocal track, these modifications had to be entered by the user. The values had to be changed and calculated depending on the song.
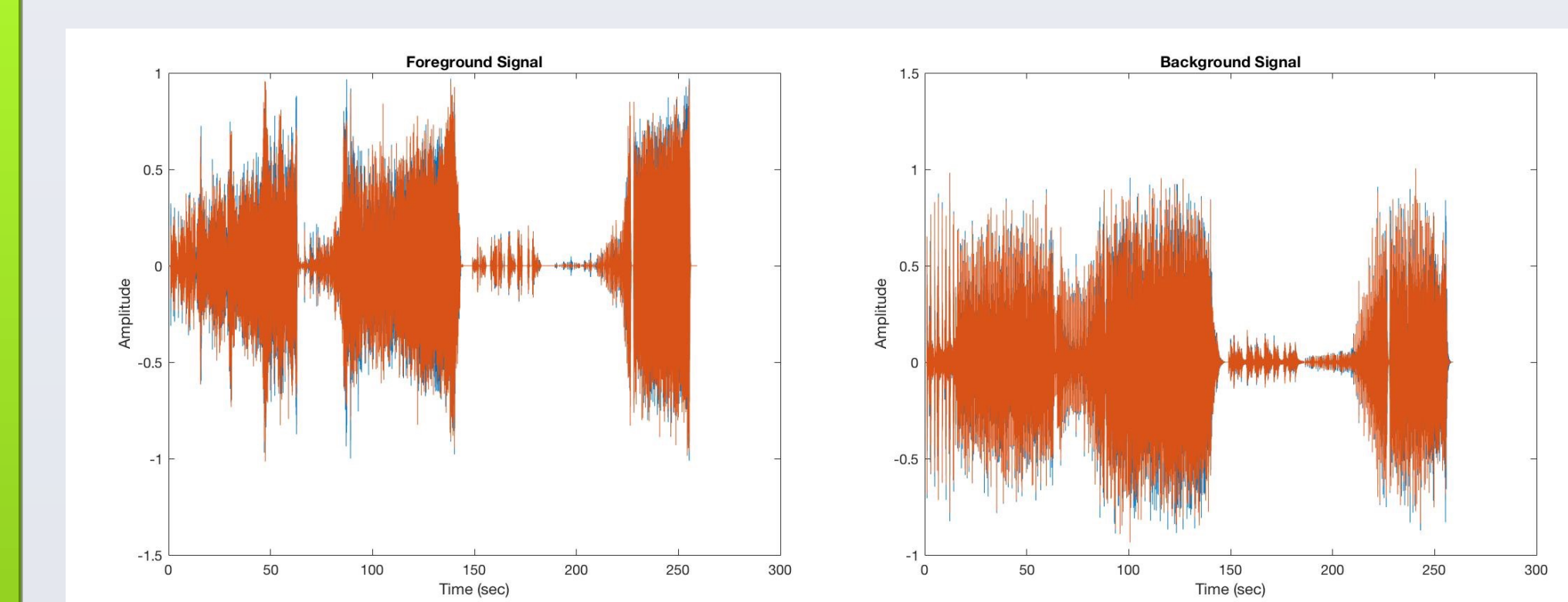
There are still some other filters that could be applied in future work. We unsuccessfully attempted to apply a Gaussian filter to remove hi-hat sounds from vocal tracks. A successful implementation of this filter may produce better results.

UNSUCCESSFUL IMPLEMENTATION OF REPET

When running REPET, songs that did not work well with the program were songs in the jazz style. We believe this is because some background instruments are placed in the foreground when they have a solo in the music. This happens very often in jazz. This style of music tends to be less repetitive than other forms of popular music, thus contributing to the unsuccessful implementation. Below are some visualizations of the attempted separation of *La La Land*'s "Someone in the Crowd"

Original REPET

Vocal Melody   *Instrumental Background*

Butterworth Filter